

## What aspects of experience can functional neuroimaging be expected to reveal?

Andrew C. Papanicolaou \*

*Division of Clinical Neuroscience-Department of Neurosurgery, The University of Texas-Houston Medical School, 1333 Moursund Suite H114, Houston, Texas 77030, USA*

Received 20 May 2006; received in revised form 24 June 2006; accepted 13 July 2006  
Available online 23 January 2007

### Abstract

The success of functional neuroimaging methods in picturing distinct cerebral activation profiles for different psychological functions has led many specialists and many more non-specialists to speculate that it will soon be possible for neuroimaging experts to sit in front of the screen of a functional brain imaging device and “read” in the changing patterns of brain activity, displayed there in real-time, what the person whose brain is imaged is experiencing from one moment to the next.

This apparently reasonable scenario presupposes that each concrete experience is associated with a distinct and unique brain activity pattern, and that these patterns are, at least in principle, discernible through functional neuroimaging. It will be argued that, for reasons of an epistemological order, even if the first assumption was true, that pattern could not be discernible, and therefore readable, even with ideal neuroimaging devices. It will also be argued that the only epistemologically, therefore, in principle, also technically feasible feat is to discern and decipher patterns of the brain activity corresponding not to concrete experiences, but to types or “kinds” of experiences, that is, to general concepts. Moreover, it will be shown that we could, in principle, discern only such patterns for the very same reason that we can know objectively only concepts, that is, the invariant features common to sets of concrete, fleeting and unrepeatably single experiences.

© 2007 Published by Elsevier B.V.

*Keywords:* Functional neuroimaging; Magnetoencephalography; Brain activity

It is fairly obvious that in our days the functional imaging methods are improving at an exceedingly rapid pace. At first they enabled us to get glimpses of brain activation patterns specific to simple sensory and motor functions. Soon after that, they revealed the outlines of the mechanisms of higher functions. These vague outlines are becoming clearer and more detailed by the day, both in terms of spatial and temporal resolution. With magnetoencephalography (MEG), for instance, the method that I am most familiar with, it is now possible to discern brain activation patterns which are interpreted as signs of the brain mechanism of functions such as language, memory or attention. These dynamic spatiotemporal activity profiles, consisting of images of several brain areas activated in sequence or in parallel, are expected to become more accurate, more complete (in terms of including more structures) and more

detailed (in terms of representing the activation of each of these structures moment by moment) especially if we combine methods such as fMRI that has an excellent spatial resolution with MEG, which has excellent temporal resolution.

These improvements give rise to reasonable questions regarding the nature of the psychological processes, the mechanisms of which may be captured in images in the near future. To be precise, at the present time, we may only interpret spatiotemporal patterns as the signs of mechanisms of functions like language or those of constituent operations of such functions, such as phonological or semantic operations. But the most poignant of these questions is this: *Will there come a time when we will be able to capture in functional images not only the mechanisms of functions, but also the brain activity that is specific to individual phenomena, which are the products of these functions.* Namely, will we be able to obtain patterns specific to, say, the perceptual experience of a rose or a human face or this specific face, or, even, this specific face encountered at this specific moment?

\* Tel.: +1 713 797 7570.

E-mail address: [apapanicolaou@uth.tmc.edu](mailto:apapanicolaou@uth.tmc.edu).

Many of us directly involved in neuroimaging and even more people not so involved in it, and therefore not intimately familiar with it, give an affirmative answer to this question. Some even go so far as to propose that there will come a time when we will be able to read the contents of the stream of consciousness of individual people in the images of their brain activity patterns as these flow on the screens of our imaging devices.

The repercussions of such a feat, if accomplished, would be tremendous. I do not even feel the need to mention its repercussions to our understanding of the brain functions. Nor do I need to mention the impact of such a feat on our ability to diagnose all sorts of psychiatric disorders. But I do need to mention, however briefly, the broader, legal, moral and political repercussions of it: the person, you see, who could read what someone thinks or feels or is about to decide, in that someone's brain activity, may not always be the diagnostician to whom that someone has voluntarily submitted his brain and his mind for examination, but it may turn out to be one of those dreaded "Big Brothers."

I suppose that the importance of the question is fairly obvious and I need not emphasize it any further. It is also obvious that it deserves direct and careful examination. It is such an examination I propose to relate to you at this point, to the best of my ability. However, having thought through it before, I have concluded that no one ought to worry about Big Brothers because no Big Brother (but, no benevolent doctor either) will ever be able to read the contents of one's consciousness no matter how far the technology improves. The only thing left to do now is to try to convince you of the soundness of this conclusion. But I propose to do a bit more than that. Namely, I will try to show which aspects of the stream of consciousness may reasonably be expected to be read as these technologies approach perfection. More analytically, I intend to argue the following.

Firstly, I will argue that the prospect of mind-reading is unrealistic, in principle. That even with perfect neuroimaging devices, with nanometer spatial resolution and nanosecond temporal resolution (if that is what you think would make them perfect), there is still no way to read the stream of human consciousness on-line or off-line.

Secondly, I will argue that this impossibility is not due to "mind" being immaterial or due to consciousness being something other than the outcome of brain processes. Whether or not mind is immaterial, whether or not consciousness is just another output of the brain has no bearing on the issue whatsoever. Rather, the impossibility is due to the same factors that make it impossible for us to know clearly and to convey to others any aspects of concrete and ever-changing experience except for that aspect that, even subjectively, is known and recognized only by reference to invariant concepts. That is, the impossibility is due to the very structure and limits of all articulate knowing.

Thirdly, I will argue that some aspects of some of the contents of consciousness, from among those that can be known, symbolically represented and communicated, may be "read" in the brain activity profiles after extensive analysis

though not on-line, that is, as they unfold in real time. Yet even those will most likely be discerned only with the full cooperation of the person whose brain is imaged.

These arguments rest, much like any and all arguments, on propositions that may or may not be acceptable and which I will now try to articulate: Needless to say, I find these propositions reasonable and I daresay you will find them reasonable as well.

The first proposition is that the experiences that constitute a person's stream of consciousness have two aspects: a repeatable aspect and a non-repeatable one. Now, the statement that each experience has two aspects—that it is unique and unrepeatable as well as repeatable—may sound enigmatic or mystical but it really is neither. Instead, it might only be a not sufficiently clear way of describing the commonest of phenomena.

For example, I have the visual experience of this pencil. Now this specific experience, of the moment just past, is as unique and unrepeatable as is the moment itself, in the context of which it occurred. At the same time, the very same visual experience is the experience of a familiar *type* or *kind* of thing, namely pencil-at-large; familiar in that, in the past, I have had several equally unique experiences of the very same *type* of object. The operative word here is "type." The unique aspects of each particular experience are surely felt (we are aware of them—otherwise they would not constitute a part of the stream of consciousness) but are ineffable. We cannot describe them because they occur only once in the entire history of the universe—therefore they are not familiar—and they are fleeting. Nevertheless, they suffice to vouch for the truth of the dictum of Heraclitus of Ephesus, the familiar one that states, "One may not wade in the same river twice" because each time it is a different stream one is wading in.

On the other hand, the repeatable aspects of the same experience, precisely because we have been exposed to them before, are both familiar and describable with words or other symbols. Those we know objectively and we can communicate our knowledge of them to others. But even so, it is only a subset of those the corresponding brain activity pattern of which we may be able to discern on the computer screens of our neuroimaging devices in the future; or this is what I intend to demonstrate.

But first, let me attempt to explain why I believe the dictum of Heraclitus (also asserted by the early philosophers of Becoming and their modern counterparts—like Henri Bergson or William James) is true; in what sense, I believe each and every experience is unique and non-repeatable: Consider two glimpses of the same object, two perceptual experiences of the same thing, say, again, this pencil. Let us in fact consider them happening one right after the other, without any third experience intervening between them, so that we can be absolutely sure that we don't forget the exact way this pencil looks to us in its two successive appearances. Now, either the two perceptions are identical or they are not. If they are identical, provided that no other experience intervened between them, it would be impossible to tell that they are two separate ones rather than one. The only way for them to be perceived as two would be to differ in something and not be exactly identical. Therefore, in claiming that they are two rather than one, we admit that we

discern a difference between them, even if we could not specify what that difference is. (No wonder we cannot, given that each of them is a fleeting, unfamiliar and unrepeatable, therefore a different experience). The possible objection that although they are identical in all respects they “seem” to be different due to their occurrence at two different times is plainly missing the mark, simply because we must have already felt them as two separate experiences, rather than one continuous experience, in order to tell that the one happened now and the other next.

So, can we say for sure that no experiential moment, no concrete experience ever repeats itself in an *identical form*? I think we can. We can, not only because our own introspection compels us to do so, but also because our reason does not afford us plausible alternatives. For one thing, the second experience in the example includes awareness of, or memory of the first. More generally, any process, be it psychological, physiological or physical, can only be real if its present phase is different than the previous one, even in that a single ion in one single synapse has changed its position, even if a single neuron has sprouted a new connection or has died in the interim. Common sense also tells us that “now” is different from “before”; that it can’t be a real “now” if it is exactly the same as the “before.”

Therefore, the question becomes: is it possible to recognize in the ever-evolving brain activity pattern pictured on the screen of even the most perfect imaging device, the sign of such unique experiences so that we may exclaim as we are watching the screen: “Now the subject had an itch on the forehead unlike any other itch he ever have had but, he thought it impolite to scratch in front of all these people, that he feels they would be ready to judge him unfairly anyway—but who cares? The itch is gone and he now enjoys the breeze off the shore and the glimpse he just had of that red rose...etc...etc...”. That’s the question.

So let me recapitulate: The answer to this question, as I said before, is *no*. The justification for that answer rests on some propositions, the first of which was: Each experience has two aspects; an aspect that repeats (and enables us to know what the experience is about) and an aspect that is ineffable, fleeting and non-repeatable that renders the entire experience unique.

Now proposition number two: To everything that goes on in consciousness plus to most of the processes that go on in the body, there corresponds a specific form of brain activation. This proposition does not require that the activity is sufficient for the corresponding conscious and the physiological processes. It does not even make the reasonable demand that the activity is necessary for them. It simply states that there is a correspondence—something that everyone, idealists included, should, I would think, accept without argument.

Acceptance of the two propositions that (1) concrete particular experiences are unique or that they have an aspect that renders them unique and non-repeatable and (2) that a particular, specific pattern of brain activity corresponds to each unique experience, leads to this inescapable conclusion: *No such unique pattern ever repeats itself*. And, due to the same reason that fleeting non-repeating unique aspects of experience cannot be recognized for what they are, it is just plain common sense that neither can their corresponding patterns be ever recognized for what they stand for.

Therefore, they can never be read; they can never be interpreted; we can never know or recognize which concrete but ineffable experiences they are the signs of.

We could stop here if my purpose were only to show that any fears that “Big Brother” may be lurking somewhere in our future, are pointless. However, that was not my only purpose. As I mentioned earlier, I also intend to discuss in this paper, what is possible for functional neuroimaging to reveal, if not patterns corresponding to the non-repeatable aspect of experiences. To that end, let us draw a second inference from proposition number two: Given that to everything that goes on in consciousness and to most physiological processes that go on simultaneously or in a temporally overlapping manner, in the body, there is a brain activity pattern unfolding in parallel, and given that the number, duration and the temporal overlap of the latter must be quite substantial (though how substantial, remains unknown) none of the corresponding brain activity patterns, buried in the great river of the global activation they themselves form, can be discerned and recognized “on-line.”

The question therefore becomes: if it is impossible to recognize the patterns corresponding to the fleeting, non-repeatable aspects of experiences neither on nor off line; and if it is similarly impossible to recognize any sign of any aspect of any experience on-line, is it at least possible to recognize and read off line (after they actually occur) the repeatable aspects of some experiences?

In the context of proposition one, it was admitted that there is indeed an aspect to some experiences that occurs again and again, and since it occurs again and again, we are able to recognize and tell (to ourselves) what these experiences are about and convey their content to others through symbols (usually words). If no aspect were repeated, there would not be any knowing whatsoever (something that has been proven beyond reasonable doubt in Plato’s dialogue “Theaetetus” 24 centuries ago). So yes, to perceive this pencil as a pencil and all instances of real or otherwise represented pencils (as e.g. in pictures) something must repeat from one experience of this particular thing to the next experience of the same thing. And that is the type, the idea, roughly the concept of which the particular percept of the thing is a token, and which idea is a sort of summary of the features common to all appearances of such tokens. Now, as you undoubtedly know, there is endless debate concerning the nature of those concepts. Plato thought that, since they endure diachronically, concepts are the only things that are real, existing, in fact, independently of the particular consciousness that may experience them. Aristotle, and I would imagine the majority of today’s psychologists and neuroscientists, think of them as mere verbal labels that we learn by exposure to specific tokens or instances—all of which have in common the set of invariant features that constitutes the concept. Fortunately, we do not need to resolve this issue in order to advance the present argument. For our purposes, it may be sufficient if we agree that (a) some aspects of experiences—the ones that allow us to tell what the experience is an experience of—repeat and (b) that (in accordance with proposition number two) they must each be associated with a corresponding brain activity pattern.

Now, can we extract from the river of global brain activity those patterns that we assume to be constituents of the global

activity and which correspond to the repeatable (and therefore potentially knowable) aspects of experiences? The answer here is yes—in principle. There are several ways in fact for doing that sort of thing, exemplified by the very familiar procedure of signal averaging: For example, to identify the pattern of brain activity corresponding to the concept of a pencil and extract that pattern from the river of the global brain activity, one simply presents to subjects the picture of a pencil repeatedly, collects and averages their brain activity contingent on the presentations of the picture and the pattern. As a result of averaging, the activity pattern present during all repetitions will emerge, whereas those patterns corresponding to all non-repeated and non-repeatable aspects of the experience will be averaged out as noise.

Then, concept by concept one may create a library of activity patterns and can use them, as we use the letters of the alphabet to possibly recognize, in the future, and guess, on their basis, the concept that may have been present in a person's consciousness during the experiment. Only this is a rather optimistic assessment because there are at least two fundamental problems that limit severely the range of concepts that can be recognized as having transpired in someone's consciousness through recognition or "reading" of their corresponding brain signs or patterns. The one is theoretical, the other practical.

Let us then face the theoretical one first: No library of brain activity patterns, corresponding to concepts, could ever be complete given the fact that new concepts arise in individuals' minds all the time. We call that "learning". We call that "progress". Given the emergence of new concepts, concepts like "cell-phone" for instance, the problem arises as to how their tokens are recognized when they appear for the first time. The empiricist answer here is, of course, "slowly over time, over many exposures to the token, we extract their common features thus we build the concept". To which answer the idealists' retort is, "how could you recognize any feature as belonging to the set that defines the concept unless you have knowledge of the concept in the first place?"

Now, as I mentioned earlier, this debate began centuries ago, yet which of the two answers to the question is the correct one is still not clear and the difficulty with building a library of activity patterns for future use still remains. And here is why.

In case the empiricist answer is correct, to obtain the activity pattern specific to any new concept, we not only need to extract the average activity pattern corresponding to 100 or 1000 tokens of it, but then we have to make sure that the average pattern is not merely due to the acoustic or the visual characteristics of the stimulus "cell-phone", but is due to the concept "cell-phone". Then we have to make sure that it differs from the brain pattern of the concept "phone-in-general", of "communication", of "wireless transmission" and so on. In short, if we really want to make sure that the pattern is specific to the concept of interest and not to any other related concept, we would have to spend entire careers computing averages. Now try to imagine going through the same process with every new concept having already gone through it for all the old ones—which are, by the way, how many?

If the empiricist solution has lost its appeal as a result of the foregoing analysis, we may want to try the idealist one. According to the idealist point of view, there are certainly new concepts but those, much like the old ones, are made of different combinations

of a *finite set of basic features* in much the same way that the host of different molecules existing in the world are merely different configurations of a few basic elements; or, in much the same way that the endless variety of possible words and sentences, is constructed by combining even smaller sets of phonemes (or graphemes). So, in order to build our library, all we have to do is to extract the brain activity patterns corresponding to the fundamental features out of which all concepts, whether they are old or new, are made. That way, we may at last have a complete library. However, there is a catch here as well: we still have to agree what are these features (is a triangle or a circle a concept or is it an irreducible feature? And what about an angle; is it a feature?) And how many features are present there?

Moreover, in the event we resolve that issue and try to find the pattern of brain activity that uniquely corresponds to each of the features we have assumed to exist, we are looking forward to the ordeal described before. Namely, the ordeal of spending entire careers in deciding first that a particular brain pattern is unique to some concept feature and to no other. Taking this into consideration, I think it would not be off the mark to assert that there are serious obstacles in ever creating a complete library of patterns corresponding even to present concepts.

Next, we have to consider the practical problem: The practical problem arises due to the two technical prerequisites of extracting concept-specific brain activity patterns, namely, repeatability of the same concept through presentation of several tokens of it, and knowledge of the precise time the concept arose in the person's brain. Both are very familiar prerequisites: in order to extract an invariant signal out of the mass of other signals in which it is embedded, that signal must be more or less invariant through its successive appearances (which requires that the corresponding conceptual or perceptual experience is also invariant across repetitions) and one should know at what point in time it occurs. These prerequisites are easier to satisfy in the case of concrete objects like a pencil or a microphone. There, you may hope to repeat perceptual experience of the same type (e.g. pencil) by simply exposing the person to the picture, of the same pencil time after time. But, evoking the same experience is much more problematic in the case of abstract concepts or function words like "and", "rejuvenation", "variety" and "however", which, even if presented in isolation or in the same context, are unlikely to evoke the same meaning each and every time, and if presented in different contexts, are almost certain to evoke different shades of meaning each time.

The practical limitation is also arising in connection to the requirement that we, the experimenters, know when the concept and therefore its corresponding brain sign arises, if we are to extract that sign properly from the global activity of the brain. This limitation is more severely felt with abstract concepts and function words that do not refer to concrete sensible objects like microphones or pencils, which we can show to subjects at a precisely known time, hoping to evoke the corresponding experience right then. And we all know that merely presenting the word "justice" for one to read in order to isolate the brain sign of the concept "justice" does not guarantee that that sign will arise in the brain (and that meaning will arise in consciousness) *at the very moment the word is presented*. Most likely, it will arise earlier

on some occasions and later on others, thus compromising the clarity and validity of the extracted average pattern.

How severe these practical limitations are, I do not pretend to know. And I daresay neither does anyone else. The reason for such a strong assertion is simple: We have yet to isolate reliable and valid patterns specific to any concept whether it be “justice” or the “red rose” such that we could, were we to be shown the functional image of roses and such, declare unequivocally and correctly: “but of course that’s the brain pattern of a rose and that one of the smell of a gardenia and that other one of the square root of number two”. We can hope, however, to estimate the severity of these practical limitations as well as the theoretical limitations discussed before, only when we have begun to accumulate a reasonably diverse library of such brain signs.

That was my argument: it leads—as far as I can tell, to the inescapable conclusion that reading even in the most detailed video of a person’s brain activity, whether “on” or “off-line”, the story that unfolds in that person’s consciousness, is not a

realistic possibility. And it is impossible to the degree that it is impossible to know, to recognize even subjectively and articulate clearly, that which constitutes the mass of elements that make each of our experiential moments unique, but also inarticulate, fleeting, inconstant, incomplete and perishable.

Nature seems to abhor Big Brothers even more than it does the proverbial vacuum. It does allow science to discern the discernible and the discernable only. That is, to discern that which abides, that which endures. In Greek, the word for truth is *aletheia*. Literally that which does not fall into oblivion right after it appears; that which abides, that which endures, that which remains invariant. That is, concepts, roughly or ideas. In principle, these are the only things we may know and for some of these we may even discover the corresponding brain signatures: the things that Plato thought constitute the very essence of reality and the very same ones his otherwise accomplished student thought of as mere words.

AUTHOR'S PERSONAL COPY